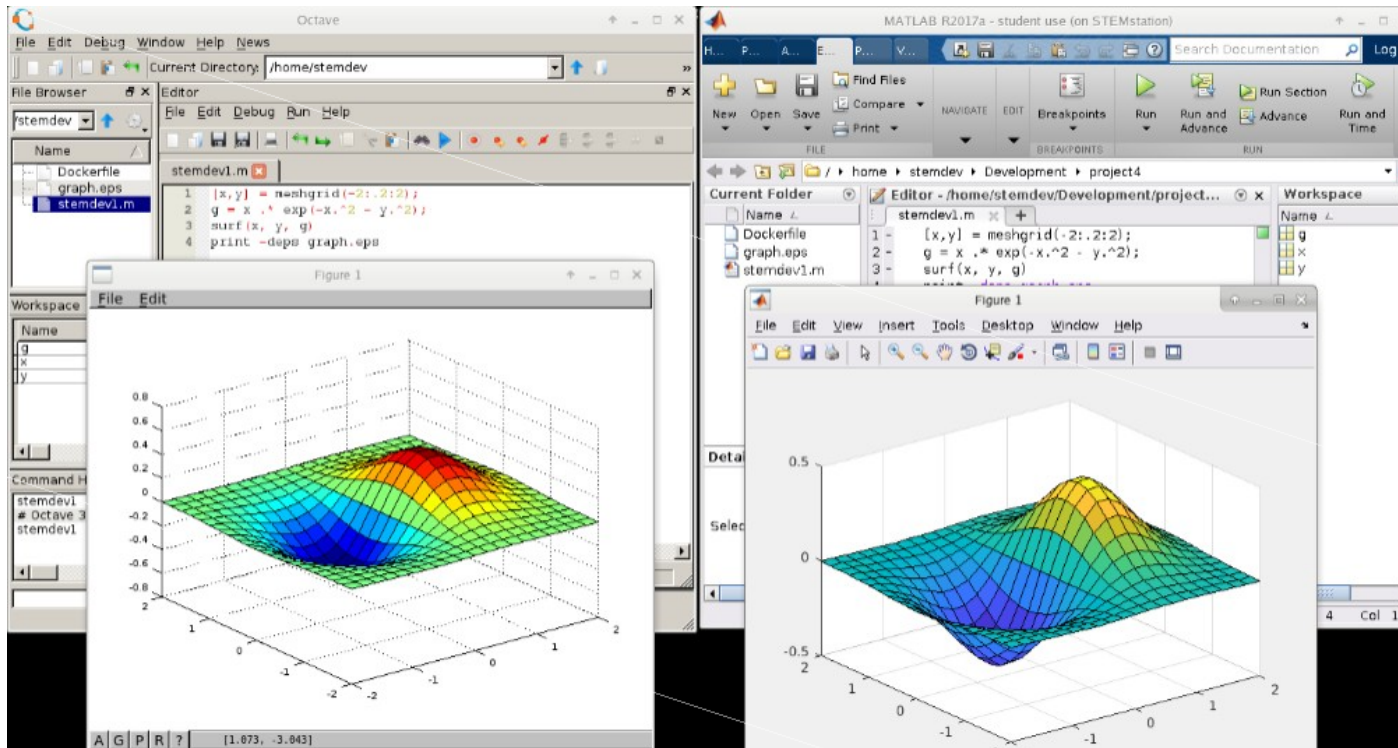


## Contenidors Linux per a entorns de Ciència i Enginyeria

**Àngel Linares Zapater**  
Enginyeria Informàtica, 2n. cicle  
Administració de Xarxes i Sistemes Operatius

**Eduard Marco Galindo**  
**Pierre Bourdin**

7 de Juny de 2017



## Contingut

- **Introducció**  
Context i justificació  
Objectius i enfocament
- **Estudi teòric**  
Els entorns STEM: classificació, rols i necessitats  
Comparativa de solucions basades en contenidors
- **Aplicació pràctica**  
Disseny i implantació d'un cas representatiu
- **Anàlisi dels resultats**  
Resultats del cas particular  
Anàlisi del cas general
- **Conclusions**  
El treball realitzat  
Treball futur

## Introducció

- Context i justificació

La recent popularitat de solucions de contenidor, com Docker en *virtualització d'aplicacions*, fa plantejar la hipòtesi de si la tècnica de contenidors pot ser també aplicable als entorns STEM.

En especial, aquesta tècnica de contenidors podria tenir un paper a jugar en la millora de la reproductibilitat dels resultats computacionals.

- Objectius i enfocament

- Identificar els rols i les necessitats computacionals en els entorns STEM
- Comparar les solucions alternatives, basades en contenidors, aplicables
- Aplicar solucions de contenidors a un cas model d'entorn STEM
- Valorar l'aplicació dels contenidors al treball computacional i com facilitaria la reproductibilitat de resultats numèrics

Per assolir aquests objectius es dissenya un cas model i implanta una aplicació, usant solucions de contenidors, per a poder valorar la seva aplicabilitat a aquest tipus d'entorns.

## Estudi Teòric

- **Classificació dels entorns STEM**

Prenen com a criteri diferencial la dimensió de la xarxa d'ordinadors necessària i el nombre d'actors implicats, es poden distingir tres classes d'entorns STEM:

- **entorn STEM individual:** escenari típic d'un enginyer/a o científic/a que treballa de forma individual i que per tant ha d'assumir tots els rols en el cicle de producció i publicació
- **entorn STEM d'equip:** escenari propi d'un grup de persones que treballa en un mateix projecte i comparteix recursos en xarxa, on alguns individus poden assumir rols especialitzats si s'escau
- **entorn STEM obert:** escenari on dos grups STEM independents intercanvien solucions i resultats computacionals, sota un paradigma productor-consumidor, els primers per a oferir una solució desenvolupada i els segons per a revisar-la.

## Estudi Teòric

- Rols en els entorns STEM

Els actors dels entorns STEM assumeixen diferents rols. Alguns actors assumeixen els diferents rols al llarg del cicle de treball i altres ho fan de forma especialitzada en un de sol, segons les necessitats concretes de cada entorn.

Es poden identificar els següents rols genèrics en un entorn STEM:

- **desenvolupador:** crea els algorismes i experiments computacionals en el seu àmbit de coneixement i/o obté uns resultats a comunicar
- **administrador:** proveeix i manega el maquinari i el programari especialitzat propi de l'entorn STEM
- **autor:** empaqueta una solució computacional i la publica per a la seva revisió per tercers
- **revisor:** obté i avalua una solució computacional publicada i sobretot, *espera poder reproduir i obtenir els mateixos resultats que l'autor però en el seu propi entorn STEM.*

## Estudi Teòric

- **Necessitats dels rols STEM**

Analitzant els rols STEM es poden inventariar una sèrie de necessitats bàsiques, pròpies d'aquest tipus d'entorn, per a cada rol.

- Aquestes necessitats es poden especificar usant *històries d'usuari*:

Rol	ID	Història d'usuari
Desenvolupadors	DEV01	Alternança de compiladors o intèrprets
	DEV02	Alternança de llibreries
	DEV03	Alternança de programari
	DEV04	Execucions en paral·lel
Administradors	ADM01	Actualització coordinada del sistema operatiu
	ADM02	Actualització coordinada de llibreries
	ADM03	Desplegament d'aplicacions al centre de càlcul
Autors	AUT01	Empaquetar aplicacions i resultats
	AUT02	Publicar aplicacions i resultats
Revisors	REV01	Obtenir aplicacions de repositoris
	REV02	Reproduir els resultats publicats

## Estudi Teòric

- Solucions alternatives als contenidors

Abans de la recent difusió i popularització dels contenidors per a la virtualització i desplegament d'aplicacions i sistemes, es podien emprar tècniques alternatives per a intentar atendre les mateixes necessitats.

Les principals alternatives a l'ús de contenidors són:

- **modificar el propi sistema amfitrió:** el sistema disponible s'adapta i reconfigura per una necessitat però en la majoria dels casos s'obtenen efectes col·laterals no desitjats i de vegades irreversibles
- **usar *chroot*:** l'operació *chroot*, per a canviar l'arrel del sistema d'arxius d'un procés UNIX, és una primera aproximació per a resoldre aquesta problemàtica i la base tècnica que donarà peu al concepte de contenidor, quan s'automatitzen i simplifiquen les passes necessàries per al seu ús
- **màquines virtuals:** alternativa que permet simular, per software, tot un nou sistema maquinari, proporcionant l'aïllament desitjat per a l'aplicació final a canvi d'una pèrdua important de rendiment entre d'altres contrapartides.

## Estudi Teòric

- Solucions basades en contenidors

Un cop es sistematitza i automatitza l'ús de l'operació *chroot*, amb totes les tasques auxiliars per a la seva correcta configuració, i a més en el nucli Linux s'implanten les crides de sistema necessàries, comencen a aparèixer les solucions de contenidors Linux.

Aquestes solucions en basen en dues característiques proporcionades pel *kernel* Linux:

- **la gestió dels espais de noms:** que permet desenvolupar eines que administren els diferents espais de noms (identificadors de processos, identificadors d'usuaris i grups d'usuaris, dispositius de xarxa, punts de muntatge, etc.) i construir les abstraccions d'aïllament de processos anomenades contenidors
- **el control de grups de processos:** que permet a aquestes solucions controlar i gestionar determinats grups de processos com un únic conjunt, de manera que tots els processos assignats a un mateix contenidor es poden manegar de forma conjunta.



## Estudi Teòric

- Solucions basades en contenidors

Les principals implantacions de contenidors Linux, tot i partir dels mateixos principis, adopten diferents especialitzacions que caldrà tenir en compte a l'hora de triar la més adient a un entorn concret.

- **LXC/LXD:** orientada a virtualitzar sistemes complets, permet crear imatges i instanciar contenidors que contenen normalment amb tot un sistema operatiu.
- **Docker:** orientada a virtualitzar aplicacions, permet crear imatges en una estructura de capes i instanciar contenidors mínims, on normalment només s'executarà el procés desitjat i les seves llibreries necessàries, de forma efímera.
- **CoreOS/rkt:** un sistema operatiu especialitzat en iniciar contenidors, usant seu propi runtime *rkt*. Seran els contenidors instanciats els que proporcionaran la funcionalitat del sistema i oferta als usuaris i a la xarxa.
- **Singularity:** especialitzada en entorns científics, on els contenidors estan adaptats als requisits típics d'aquestes aplicacions STEM, orientades a usar al màxim els recursos disponibles en l'amfitrió.
- **runV:** solució que proporciona un nucli Linux diferent a cadascun dels contenidors iniciats.
- **shifter:** que permet, entre altres funcionalitats, partir de les imatges de contenidors per a fer-hi una operació *chroot* simple.

## Aplicació Pràctica

- Disseny i implantació d'un cas representatiu

Per tal de validar l'aplicabilitat de les solucions basades en contenidors en els entorns STEM es dissenya un cas model que representa el cicle bàsic de producció, publicació i revisió de solucions.

Es creen i configuren els següents components:

- **estació de treball Debian:** on s'hi instal·la el *runtime* de contenidors Docker per a validar les històries d'usuari dels desenvolupadors i les dels autors.
- **servidor Ubuntu Server:** on s'hi instal·la la solució de contenidors LXC/LDX per tal de validar les històries d'usuari dels administradors.
- **repositori DockerHub:** per a publicar les imatges Docker per part dels autors i permetre la cerca i descàrrega de les imatges per part dels revisors.
- **clients pels revisors:** on s'hi instal·la el *runtime* de Docker per a descarregar i executar les imatges publicades al repositori. Es comprova com es pot fer amb:
  - client Windows
  - client macOS
  - màquina virtual Linux

## Anàlisi dels Resultats

- Resultats del cas particular

Un cop implantant el model en una aplicació pràctica i validades totes les històries d'usuari identificades es comprova que totes compleixen els seus respectius criteris d'acceptació prèviament definits.

Es comprova que per a cada rol identificat es compleix que:

- **desenvolupadors:** poden crear i usar contenidors per a les seves necessitats d'alternar compiladors, llibreries i aplicacions d'alt nivell, així com crear múltiples instàncies concurrents de contenidors.
- **administradors:** poden proveir i mantenir llibreries d'imatges de contenidors per la seva xarxa i mantenir de forma coordinada les imatges dels contenidors. També poden delegar als desenvolupadors i autors part de la gestió d'aquestes imatges.
- **autors:** poden empaquetar solucions, incloent codi objecte, codi font, dades i documentació i publicar aquestes imatges en el repositori pels oferir-los als revisors.
- **revisors:** poden descarregar i executar les imatges disponibles al repositori de contenidors Linux, fins i tot si el seu entorn no disposa de clients amb aquest sistema operatiu.

## Anàlisi dels Resultats

- Valoració del cas general

El cas particular analitzat permet extrapolar els resultats al cas general i valorar l'aplicabilitat de les solucions basades en contenidors en els entorns STEM.

Les principals conclusions són:

- en la mesura que el cicle de producció i publicació de solucions computacionals dels entorns STEM és similar a l'àmbit de desenvolupament de software, les tècniques del darrer es poden aplicar al primer.
- el ritme d'adopció en els entorns STEM de noves solucions i noves tècniques com els contenidors no és, ni ha de ser, el mateix que en l'àmbit del software.
- la promesa d'execució universal de solucions computacionals, implícita en la hipòtesi de l'aplicabilitat dels contenidors Linux als entorns STEM, no es encara completa per a tots els entorns informàtics i arquitectures de computadors possibles.
- el punt més feble de les solucions de contenidors en els entorns STEM és precisament la seva principal característica, que resulta beneficiosa en altres entorns: el nucli Linux és el mateix per a tots els processos siguin els nadius en el sistema amfitrió o els que, aparentment, són independents en el contenidor.

## Anàlisi dels Resultats

- **Contenidors Linux i reproductibilitat computacional**

Els contenidors Linux passaran a formar part de l'inventari d'eines informàtiques en els entorns STEM i representen un avenç vers la reproductibilitat científica.

En aquest sentit les dues conclusions principals sobre l'aplicabilitat dels contenidors a la reproductibilitat computacional són:

- proporcionen tècniques estandarditzades d'alt nivell per a l'empaquetat i la distribució de solucions i experiments computacionals.
- no proporcionen independència de l'arquitectura del processador físic, la qual cosa pot tenir implicacions en aquells entorns STEM on sigui important una gran precisió numèrica.

## Conclusions

- **Sobre el treball realitzat**

- El treball realitzat ha permès contrastar la hipòtesi inicial sobre la possible aplicació als entorns STEM de les solucions de contenidors que s'han popularitzat en l'àmbit de l'enginyeria del software.
- Aquestes solucions han estat prou simples d'implantar (d'aquí la seva recent popularitat i difusió) i d'aplicar als casos bàsics identificats per a aquests entorns (en tasques similars, sinó idèntiques, a les de desenvolupament de programari).
- En un entorn STEM concret i real, una aproximació més realista seria de tipus iteratiu adoptant progressivament aquestes tècniques tant pel desenvolupadors com pels administradors.

- **Treball futur**

- El treball realitzat es pot ampliar en el sentit d'aprofundir en casos més concrets i realistes que els vistos en el model plantejat. Això correspondria a fer noves iteracions sobre aquest model inicial plantejat.
- A mida que s'usin més les solucions de contenidors en els entorns STEM es faran cada cop més paleses les necessitats específiques i els requisits propis d'aquests entorns STEM respecte el desenvolupament general de software, que caldria analitzar i elaborar amb més profunditat.